AD-A204 890

DTIC FILE COPY

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Robust Estimation of Mixture Constants		5. TYPE OF REPORT & PERIOD COVERED Reprint
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Douglas J. Warren and John B. Thomas		8. CONTRACT OR GRANT NUMBER(s) N00014-81-K-0146
9. PERFORMING ORGANIZATION NAME AND ADDRESS Information Sciences and Systems Laboratory, Department of Electrical Engineering, Princeton University, Princeton, NJ 08544		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR SRO-103
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research (Code 411SP), Department of the Navy, Arlington, VA 22217		12. REPORT DATE March 1987
		13. NUMBER OF PAGES 7
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

DTIC
SELECTE
FEB 2 8 1989
S D
D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Appears in the Proceedings, 1987 Conference on Information Sciences and Systems, The John Hopkins University, Baltimore, MD.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

robust estimation ;
mixture density ;
M-estimators . (jhd)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The problem of estimating the mixture constant for a mixture process is investigated. The asymptotic properties of M-Estimators are discussed. Conditions are given under which such estimators are asymptotically normal and consistent. It is shown that the maximum likelihood estimate and estimates based on moments satisfy these conditions. For instances in which there is undertainty about the distributions composing the mixture distribution the robust estimator is found to be a censored version of the nominal MLE nonlinearity. Some numerical results on the existence of robust solutions are presented.

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

PROCEEDINGS

OF

THE 1987 CONFERENCE

ON

INFORMATION SCIENCES AND SYSTEMS

PAPERS PRESENTED

March 25, 26 and 27

PROGRAM DIRECTORS

Wilson J. Rugh
Brian L. Hughes

Department of Electrical Engineering

The Johns Hopkins University

Baltimore, Maryland 21218

ROBUST ESTIMATION OF MIXTURE CONSTANTS

Douglas J. Warren

Department of Electrical Engineering and Computer Science
University of Illinois at Chicago

John B. Thomas

Department of Electrical Engineering
Princeton University

ABSTRACT

The problem of estimating the mixture constant for a mixture process is investigated. The asymptotic properties of M-Estimators are discussed. Conditions are given under which such estimators are asymptotically normal and consistent. It is shown that the maximum likelihood estimate and estimates based on moments satisfy these conditions. For instances in which there is uncertainty about the distributions composing the mixture distribution the robust estimator is found to be a censored version of the nominal MLE nonlinearity. Some numerical results on the existence of robust solutions are presented.

§1 Introduction

This paper will investigate the problem of estimating the constant ε from observations of a noise process with the distribution

$$\mu = \mu_0 + \varepsilon\nu \tag{1.1}$$

where μ_0 is the known nominal distribution and ν is a known signed measure with total variation 0. Clearly, the choice of ν is constrained by the fact that μ is a distribution function. If $\nu = \mu_1 - \mu_0$ then we have

$$\mu = (1-\varepsilon)\mu_0 + \varepsilon\mu_1 \tag{1.2}$$

μ has the form of the traditional mixture distribution.

Knowledge of ε is useful for at least two reasons. Firstly, and most generally, it gives us an increased level of knowledge about the actual distribution of the noise. Secondly, certain noise environments contain bursts of high energy impulse noise superimposed on a low level background noise[5,6]. In such a situation, the noise may be thought of as alternating between two states. If, in the low energy state the noise distribution is μ_0 and in the impulsive state the distribution is μ_1, then ε is the fraction of time the noise is in the high energy state. In a very similar manner, estimates of ε may also be used as estimates of the prior distribution of any dual distribution problem. In such a case we observe the noise over a long period of time and model it as a mixture process. The resulting estimate of ε is an estimate of the prior probability of μ_1.

In this paper we will discuss both the general problem of estimating the mixture constant and the robust problem. In §2 we discuss the asymptotic properties of M-estimators for ε, including consistency and asymptotic normality. As would be expected, we find that the asymptotic variance is minimized by the maximum likelihood estimator. Some M-estimators which can be used to estimate ε are discussed briefly in §3. In approaching the robust problem we assume that there is some uncertainty as to the exact forms of both the contaminant ν and the nominal μ_0. In §4 we find, within certain constraints, the worst distribution μ_0, and the worst contaminant ν, where by worst we mean producing the highest asymptotic variance. §5 contains a summary of the preceding results.

§2 Estimation of the Mixture Constant

The problem we are interested in is as follows. We are given the observation sequence $\mathbf{x}=(x_1,x_2,\ldots,x_n)$ which is a realization of the sequence of real-valued random variables $\mathbf{X}=(X_1,X_2,\ldots,X_n)$. The $\{X_i\}$ are independent, identically distributed and have the distribution given by (1.1). Our goal is to estimate ε from $\mathbf{x}$. Throughout this paper we will assume that μ is a strictly increasing distribution. Furthermore, in the interest of mathematical tractability and because it includes the maximum likelihood estimate (MLE), we will confine this discussion to the class of M-estimators for ε.

By an M-estimator for ε we mean any member ξ_n of the set of points E which minimize the functional

$$Q(\xi,\mathbf{x}) = \sum_{i=1}^{n}\rho(\xi,x_i)$$

so that

$$\min_{\xi} Q(\xi,\mathbf{x}) = Q(\xi_n,\mathbf{x}) \tag{2.1}$$

where ρ is continuous and convex in its first argument and $\xi_n \in E$. If

$$\eta(\xi,x) = \frac{\partial}{\partial\xi}\rho(\xi,x)$$

is continuous then

$$\sum_{i=1}^{n}\eta(\ell_n, x) = 0 \tag{2.2}$$

is an equivalent condition to that of Eq. (2.1). Since ρ is convex in ξ, η is non-decreasing in ξ; if ρ is strictly convex then η will be monotonically increasing.

Asymptotic Properties of M-estimators

For the remainder of this section we will investigate the asymptotic properties of ℓ_n. Facts 1-4 are taken from [1] and are included for thoroughness.

Fact 1. *Q is convex and E is non-empty, convex and compact. For strictly convex ρ, E contains a single point.*

Fact 2. *Let $\lambda(\xi) = \int\eta(\xi,x)d\mu(x)$ and assume there is a ξ_0 such that $\lambda(\xi_0)$ exists and is finite.*

1) *Then $\lambda(\xi)$ exists for all ξ (but is not necessarily finite).*

2) *Let*

$$S = \{x: \frac{\partial}{\partial\xi}\eta(\xi,x) > 0 \ a.e.(\mathrm{m})\}$$

where m indicates Lebesgue measure. Assume $\mu(S) > 0$ and that if $x \in \neg S$ then $\eta(\xi,x) = 0$ a.e. (m). Here $\neg S$ indicates the complement of S. If i) $\rho(\xi,x) \to +\infty$ as $\xi \to \pm\infty$ for all $x \in S$ or ii) for some ξ_*, $\lambda(\xi_*) = 0$ then iii) $\lambda(\xi)$ is monotonically increasing and is strictly positive (negative) for large positive (negative) values of ξ.

Proof. See [1] and [8].

Fact 3 (Consistency). *Assume that there is some ϵ for which $\lambda(\xi) < 0$ for all $\xi < \epsilon$ and $\lambda(\xi) > 0$ for all $\xi > \epsilon$. Then $\ell_n \to \epsilon$ almost surely and in probability.*

Fact 4 (Asymptotic Normality). *Assume*

1) $\lambda(\epsilon) = 0$,

2) $\lambda(\xi)$ is differentiable at $\xi = \epsilon$ and $\lambda'(\epsilon) > 0$,

3) $\int\eta^2(\xi,x)d\mu(x)$ is finite and continuous at $\xi = \epsilon$.

Then $n^{1/2}(\ell_n - \epsilon)$ is asymptotically normal with asymptotic mean zero and asymptotic variance

$$\sigma_\mu^2(\eta) = \frac{\int\eta^2(\epsilon,x)d\mu(x)}{[\lambda'(\epsilon)]^2} \tag{2.3}$$

where

$$\lambda'(\epsilon) = [\frac{\partial}{\partial\xi}\int\eta(\xi,x)d\mu(x)]\,|_{\xi=\epsilon} = \int[\frac{\partial}{\partial\epsilon}\eta(\epsilon,x)]d\mu(x) \tag{2.4}$$

Proof. See [1] and [8].

Minimization of Asymptotic Variance

Using the chain rule and Eq. (1.1) and recalling that $\lambda(\epsilon) = 0$, we can rewrite (2.4) as

$$\lambda'(\epsilon) = \frac{d}{d\epsilon}\left[\int\eta(\epsilon,x)d\mu(x)\right] - \int\eta(\epsilon,x)d\nu(x) \tag{2.5}$$

$$= -\int\eta(\epsilon,x)d\nu(x)$$

Substituting (1.1) and (2.5) in (2.3) yields

$$\sigma_\mu^2(\eta) = \frac{\int\eta^2(\epsilon,x)[d\mu_0(x) + \epsilon d\nu(x)]}{\left[\int\eta(\epsilon,x)\frac{-d\nu(x)}{d\mu_0(x) + \epsilon d\nu(x)}[d\mu_0(x) + \epsilon d\nu(x)]\right]^2} \tag{2.6}$$

By the Cauchy-Schwarz inequality

$$\sigma_\mu^2(\eta) \ge \left\{\int\left[\frac{-d\nu(x)}{d\mu_0(x) + \epsilon d\nu(x)}\right]^2[d\mu_0(x) + \epsilon d\nu(x)]\right\}^{-1} \tag{2.7}$$

where equality holds if and only if $\eta(\xi,\cdot) = c\,d\nu(\cdot)/[d\mu_0(\cdot) + \xi d\nu(\cdot)]$.

Assume that both μ_0 and ν are differentiable. Let ρ_0 be the penalty function for the maximum likelihood estimate of ϵ. Then

$$\rho_0(\xi,x) = -\ln[\frac{d}{dx}\mu_0(x) + \xi\frac{d}{dx}\nu(x)]$$

and

$$\eta_0(\xi,x) = \frac{\partial}{\partial\xi}\rho_0(\xi,x) = \frac{-d\nu(x)}{d\mu_0(x) + \xi d\nu(x)} \tag{2.8}$$

Then from Eqs. (2.6)-(2.8) we have

$$\sigma_\mu^2(\eta_0) = \min_\eta \sigma_\mu^2(\eta).$$

Hence it is seen that the MLE has the smallest asymptotic variance amongst all M-estimators.

§3 Examples of M-Estimators for ϵ

In the preceding section we discussed the asymptotic properties of M-estimators for ϵ based upon observations from a process having the distribution $\mu = \mu_0 + \epsilon\nu$. We now investigate some examples of these estimators. The problem here is somewhat different than the location estimation problem and as would be expected we cannot define our penalty functions in precisely the same manner. For example, if $\hat{m}_n$ is the minimum squared error estimate of location, it is chosen to be that ξ which minimizes the l_2 norm of the difference between x, the sequence of observations, and $(\xi, \xi, \ldots, \xi)$. The analogous procedure for estimating ϵ would be to minimize the l_2 norm between the sequences $\mathbf{f} = (f(x_1), f(x_2), \ldots, f(x_n))$ and $\mathbf{g} = (g(\xi,x_1), g(\xi,x_2), \ldots, g(\xi,x_n))$. Since estimating ϵ can be viewed as a problem of density estimation, it seems logical to pick f and g so that they are both estimates of the density associated with μ. However, unless f actually *is* the density of μ, $f(x_i)$ will be biased (since it is based on only one sample) and minimizing $\|\mathbf{f} - \mathbf{g}\|$ will produce an inconsistent estimate of ϵ. Hence we would have to use

$$Q(\xi,x) = \sum_{i=1}^{n}\rho_n(\xi,x_i) = \sum_{i=1}^{n}\left[\frac{d\mu_0(x_i)}{dx_i} + \xi\frac{d\nu(x_i)}{dx_i} - f_n(x,x_i)\right]^2$$

where $f_n(x,x_i)$ is an estimate of the density of μ based upon x and evaluated at the point x_i. We now have a penalty function whose form (rather than simply the number of terms in the sum) depends implicitly on n. Hence, the resulting ℓ_n is not an M-estimator. Note that

$E_\mu\, \rho_n(\epsilon,x)$ is the mean squared error in the density estimate $\hat{f}_n(x,x)$. If the estimate is consistent then the mean squared error tends to zero and, since ρ is positive, $\rho \to 0$ *a.e.* (μ).

This example demonstrates the necessity for care in choosing penalty functions for M-estimators for ϵ. We must also keep in mind the constraints put on ρ in order to ensure that the resulting estimate is consistent and asymptotically normal. These constraints are given in §2.

We now look at some examples of M-estimators for ϵ; specifically, we will investigate the maximum likelihood estimate and estimators based on the squared error of a moment or moments.

The maximum likelihood estimate

As usual the maximum likelihood estimate is defined as that ξ which minimizes the functional

$$Q(\xi,x) = \sum_{i=1}^{n} \rho_0(\xi,x_i) = \sum_{i=1}^{n} -\ln\left[\frac{d}{dx}\mu_0(x) + \xi\frac{d}{dx}\nu(x)\right] \quad (3.1)$$

Clearly, the individual terms of the sum in (3.1) are defined only for those ξ for which the bracketed expression is positive. Hence the sum exists only for those ξ for which *all* the bracketed terms are positive. Using this fact alone we can prove that, under certain restrictions, if $\lim \xi_n$ exists then it lies in $[0,\epsilon]$.

Proposition 1. *If μ is absolutely continuous with respect to Lebesgue measure, $d\nu/d\mu_0$ is continuous, $\sup d\nu/d\mu_0 = \infty$ and $\inf d\nu/d\mu_0 = -1/\epsilon$ then $\overline{\lim}\, \xi_n \le \epsilon$ and $\underline{\lim}\, \xi_n \ge 0$.*

Proof. See [8].

What is remarkable here is that we have shown that the estimate is confined to the interval $[0,\epsilon]$ without even considering the minimization of (3.1). If we weaken the restriction on $d\nu/d\mu_0$ so that its range is required to be $(-1,\infty)$, then the conclusion is that $\overline{\lim}\, \xi_n \le 1$ and $\underline{\lim}\, \xi_n \ge 0$.

Proposition 2. *If $|d\nu(x)/dx|$ is positive on a set of positive μ-probability, then the maximum likelihood estimator ξ_n is strongly consistent and $n^{1/2}(\xi_n - \epsilon)$ is asymptotically normal with asymptotic mean 0 and asymptotic variance*

$$\sigma_\epsilon^2(\eta_0) = \left\{\int\left[\frac{-d\nu(x)}{d\mu(x)}\right]^2 d\mu(x)\right\}^{-1}$$

where $\eta_0(\xi,x)$ is given by Eq. (2.8).

Proof. See [8].

Squared error in moments estimates

We can define penalty functions for M-estimators in terms of the known moments of μ_0 and ν (where by the ith moment of ν we mean $\int x^i d\nu$). The general form of such a penalty function is:

$$\rho_m(\xi,x) = \sum_{i=1}^{m} a_i\,[\overline{m_i} + \xi\overline{n_i} - x^i]^2$$

where $\overline{m_i}$ and $\overline{n_i}$ are the ith moments of μ_0 and ν, respectively, and a_i is a non-negative weighting factor.

Proposition 3. *If $\overline{n_i}$ is non-zero for any i between 1 and m and both $\overline{n_i}$ and $\overline{m_i}$ are finite for i between 1 and $2m$, then the squared moment error estimate ξ_{m_n}, based on $\rho_m(\xi,x)$ is strongly consistent and $n^{1/2}(\xi_{m_n} - \epsilon)$ is asymptotically normal with asymptotic mean 0 and asymptotic variance given by (2.3).*

Proof. See [8].

§4 Robust Estimation of ϵ

In this section we will investigate the robust procedure for estimating ϵ from observations from a process having the distribution

$$\mu = (1-\epsilon)\mu_0 + \epsilon\mu_1 \quad (4.1)$$

where both μ_0 and μ_1 are probability measures. This is a specific case of the form for μ we have discussed in the preceding sections; that is, $\nu = \mu_1 - \mu_0$. The corresponding nonlinearity for the MLE for ϵ is

$$\eta_0(\xi,x) = \frac{-d\nu(x)}{d\mu_0(x) + \xi d\nu(x)} = \frac{d\mu_0(x) - d\mu_1(x)}{(1-\xi)d\mu_0(x) + \xi d\mu_1(x)}$$

We will assume that both μ_0 and μ_1 are mixtures of a known nominal probability measure and an unknown contaminant probability measure. We then have

$$\mu_0 = (1-\alpha_0)\mu_{0n} + \alpha_0\mu_{0c} \quad (4.2)$$

and

$$\mu_1 = (1-\alpha_1)\mu_{1n} + \alpha_1\mu_{1c} \quad (4.3)$$

where α_i, $i=1,2$ are known constants, μ_{in}, $i=1,2$ are known distributions and μ_{ic}, $i=1,2$ are unknown distributions. We will also require that the nominal value of $\eta_0(\xi,x)$,

$$\eta_{0n}(\xi,x) = \frac{d\mu_{0n}(x) - d\mu_{1n}(x)}{(1-\xi)\mu_{0n}(x) + \xi\mu_{1n}(x)} \quad (4.4)$$

be a monotonic (either non-decreasing or non-increasing) function of x. Note that if we let $l_n = d\mu_{0n}/d\mu_{1n}$ then we can rewrite (4.4) as

$$\eta_{0n} = \frac{l_n - 1}{(1-\xi)l_n + \xi}. \quad (4.5)$$

Differentiation of (4.5) with respect to l_n demonstrates that η_{0n} is a monotonic function of the likelihood ratio l_n. Hence monotonicity of the two functions is equivalent and we can restrict ourselves to those pairs of nominal distributions for which the likelihood ratio is monotonic.

For the remainder of this section we will restrict the discussion to those μ_{0n} and μ_{1n} which are continuous, absolutely continuous with respect to Lebesgue measure and for which l_n is continuous with range $(0,\infty)$. We will show that if the monotonicity condition for the nominal distributions μ_{0n} and μ_{1n} is satisfied, then, under certain conditions, there exists a distribution μ_* which maximizes the variance of the MLE for ϵ within the set of distributions formed from those nominals. More formally, let

$$C = \{\mu: \mu = (1-\epsilon)[(1-\alpha_0)\mu_{0n} + \alpha_0\mu_{0c}] + \epsilon[(1-\alpha_1)\mu_{1n} + \alpha_1\mu_{1c}]\}$$

where μ_{0n} and μ_{1n} are fixed, $d\mu_{0n}/d\mu_{1n}$ is monotonic and

μ_{0b} and μ_{1b} are any distributions. The worst distribution in C is written as

$$\mu_* = (1-\varepsilon)\mu_{0*} + \varepsilon\mu_{1*}$$

where

$$\mu_{0*} = (1-\alpha_0)\mu_{0s} + \alpha_0\mu_{0b*}$$

and

$$\mu_{1*} = (1-\alpha_1)\mu_{1s} + \alpha_1\mu_{1b*}$$

Using the minimax criterion, μ_* satisfies

$$\min_{\eta \in H} \sigma_{\mu_*}(\eta) = \max_{\mu \in C} \sigma_{\mu}(\eta_*)$$

where H is the set of all M-estimator nonlinearities and

$$\eta_*(\xi) = \frac{d\mu_{0*} - d\mu_{1*}}{(1-\xi)d\mu_{0*} + \xi d\mu_{1*}}.$$

is the nonlinearity for the MLE of ε when the noise has the distribution μ_*.

As is by now expected, η_* will turn out to be very similar to a censored version of the MLE nonlinearity for noise having the distribution

$$\mu_s = (1-\varepsilon)\mu_{0s} + \varepsilon\mu_{1s}.$$

The technique used to construct the worst distributions μ_{0*} and μ_{1*} is very much like that used in [2] and [5] to find the worst pair of distributions for risk for hypothesis testing. For that problem, the processor nonlinearity was a likelihood ratio, whereas here, our nonlinearity is a monotonl function of the likelihood ratio of μ_0 and μ_1. In either case, the resulting function is a censored version of the nominal nonlinearity.

The asymptotic variance of an M-estimator for ε can be written as (see (2.3)-(2.5)):

$$\sigma_\mu^2 = \frac{\int \eta^2(\varepsilon)[(1-\varepsilon)d\mu_0 + \varepsilon d\mu_1]}{\left[\int \eta'(\varepsilon)[(1-\varepsilon)d\mu_0 + \varepsilon d\mu_1]\right]^2} \qquad (4.6)$$

where for simplicity we have written $\eta(\varepsilon)$ for $\eta(\varepsilon,x)$ and $\eta'(\varepsilon)$ for $\partial/\partial\varepsilon\,[\eta(\varepsilon,x)]$. Let

$$\eta_1(\xi) = \frac{(1-\alpha_0)d\mu_{0s} - (1-\alpha_1)d\mu_{1s}}{(1-\xi)(1-\alpha_0)d\mu_{0s} + \xi(1-\alpha_1)d\mu_{1s}} \qquad (4.7)$$

Note that η_1 is monotonic in ℓ_s. Assume $\eta(\xi)$ is a symmetrically censored version of $\eta_1(\xi)$; for example,

$$\eta(\xi) = \begin{cases} -k, & \eta_1 \le -k \\ \eta_1(\xi), & -k < \eta_1 < k \\ +k, & \eta_1 \ge k \end{cases} \qquad (4.8)$$

where k is a positive number. Then by the monotonicity of the likelihood ratio ℓ_s, η is a monotonic function with

$$\eta'(\xi) = \begin{cases} 0, & \eta_{1s} \le -k \\ \eta_1'(\xi), & -k < \eta_{1s} < k \\ 0, & \eta_{1s} \ge k \end{cases}$$

If μ_0 and μ_1 are given by (4.2)-(4.3), then the portion of the denominator of (4.6) which is effected by the choice of μ_{0b} and μ_{1b} is

$$d = \int \eta'(\varepsilon)[(1-\varepsilon)\alpha_0 d\mu_{0b} + \varepsilon\alpha_1 d\mu_{1b}]^2$$

and the portion of the numerator of (4.6) which is likewise effected is

$$n = \int \eta^2(\varepsilon)[(1-\varepsilon)\alpha_0 d\mu_{0b} + \varepsilon\alpha_1 d\mu_{1b}].$$

Since η is monotonic, the value of d which minimizes the denominator of (4.6) is zero. Similarly, since the maximum value of η^2 is k^2, the value of n which maximizes the numerator of (4.6) is $[(1-\varepsilon)\alpha_0 + \varepsilon\alpha_1]k^2$. Hence, in order to maximize $\sigma_\mu^2(\eta)$ we would choose μ_{0b} and μ_{1b} so that $d\mu_{0b}$ and $d\mu_{1b}$ are zero except when $|\eta_1| \ge k$. If μ_0 and μ_1 can be found for which this restriction on the support of the contaminant is satisfied and for which η, as given by (4.8), is the MLE nonlinearity, then we have solved the minimax problem.

For the hypothesis testing problem described in [2] the worst pair of distributions for discrimination were, in some sense, as close to each other as possible. For the estimation of the mixture constant, the result is similar. Clearly, if μ_0 and μ_1 are the same distribution, it will be very difficult to estimate ε. In such a case, since η_0 is zero for all x, we have censored η_1 at zero. The resulting nonlinearity is one way of saying that nothing is known and nothing can be learned.

The general robust solution

The general form for the solution for the robust problem is given by:

$$\mu_{0*}' = \begin{cases} (1-\alpha_0)c_1\mu_{0s}', & \ell_s \ge d_2 \\ (1-\alpha_0)\mu_{0s}', & d_1 < \ell_s < d_2 \\ (1-\alpha_0)d_1 c_2\mu_{1s}', & \ell_s \le d_1 \end{cases} \qquad (4.9)$$

$$\mu_{1*}' = \begin{cases} (1-\alpha_1)\dfrac{c_1}{d_2}\mu_{0s}', & \ell_s \ge d_2 \\ (1-\alpha_1)\mu_{1s}', & d_1 < \ell_s < d_2 \\ (1-\alpha_1)c_2\mu_{1s}', & \ell_s \le d_1 \end{cases} \qquad (4.10)$$

where $A = (c_1, c_2, d_1, d_2)$ is a non-unique set of positive constants. If

$$d_2 = \frac{r[2\varepsilon(r-d_1) + d_1]}{2\varepsilon(r-d_1) + 2d_1 - r} \qquad (4.11)$$

where $r = (1-\alpha_1)/(1-\alpha_0)$ then η_* is a symmetrically censored version of η_1 as in (4.8). We can also censor on one side only. Rewriting (4.7) as

$$\eta_1 = \frac{\frac{1-\alpha_0}{1-\alpha_1}\frac{d\mu_{0e}}{d\mu_{1e}} - 1}{(1-\varepsilon)\frac{1-\alpha_0}{1-\alpha_1}\frac{d\mu_{0e}}{d\mu_{1e}} + \varepsilon} \tag{4.12}$$

makes it easy to see that the range of η_1 is $\{-1/\varepsilon, 1/(1-\varepsilon)\}$. If, for example, $\varepsilon < 1/2$ then we can censor η_1 at $-k$ where $1/(1-\varepsilon) < k < 1/\varepsilon$. Clearly, we have not effected the value of η_1 where it is positive while at the same time we have censored so that maximum magnitude and (assuming without loss of generality that l_e is monotonically increasing) minimum slope are attained on the same set.

Since $\mu_{0c} = (1-\alpha_0)\mu_{0e} + \alpha_0\mu_{0h}$, we must have $\mu_{0c} - (1-\alpha_0)\mu_{0e} \geq 0$. Similarly, we must have $\mu_{1h} = \mu_{1c} - (1-\alpha_1)\mu_{1e} \geq 0$. Inspection of (4.9)-(4.10) reveals that this requirement is equivalent to requiring that c_1 and c_2 are both greater than 1. Furthermore, for μ_{0c} and μ_{1c} to be distributions, they must have total mass of unity. Hence, we must choose A with d_2 (for symmetric censoring) given by Eq. (4.11), $c_i \geq 1$, $i=1,2$, and

$$1 = (1-\alpha_0)[c_1\mu_{0e}(l_e \geq d_2) + \mu_{0e}(d_1 < l_e < d_2) + d_1c_2\mu_{1e}(l_e \leq d_1)] \tag{4.13}$$

$$1 = (1-\alpha_1)[\frac{c_1}{d_2}\mu_{0e}(l_e \geq d_2) + \mu_{1e}(d_1 < l_e < d_2) + c_2\mu_{1e}(l_e \leq d_1)] \tag{4.14}$$

With $c_i = 1$, $i=1,2$, (3.4.25) and (3.4.26) become

$$(1-\alpha_0)[\mu_{0e}(l_e > d_1) + d_1\mu_{1e}(l_e \leq d_1)] = 1 \tag{4.15}$$

and

$$(1-\alpha_1)[\frac{1}{d_2}\mu_{0e}(l_e \geq d_2) + \mu_{1e}(l_e \leq d_2)] = 1. \tag{4.16}$$

It is shown in [2] that there always exist d_1' and d_2' for which (4.15) and (4.16) are true. Unfortunately, in general, this solution will not satisfy (4.11). Let the bracketed terms in (4.13)-(4.16) be known as $f(A)$, $g(A)$, $f(d_1)$, and $g(d_2)$, respectively. Note that both $f(A)$ and $g(A)$ are increasing functions of c_1 and c_2. Furthermore, as shown in [2], $f(d_1)$ is an increasing function of d_1 and $g(d_2)$ is a decreasing function of d_2. Then if $A = (c_1, c_2, d_1, d_2)$ is a solution of (4.13) and (4.14), we must have $d_1 \leq d_1'$ and $d_2 \geq d_2'$. This fact will lead to a precondition for the use of symmetric censoring. Let $k = \eta_1|_{l_e=d_1}$. Then from (4.11) and (4.12) it can be shown that $\eta_1|_{l_e=d_1} = -k$,

$$d_1 = r\,\frac{1-\varepsilon k}{1+(1-\varepsilon)k} \tag{4.17}$$

and

$$d_2 = r\,\frac{1+\varepsilon k}{1-(1-\varepsilon)k} \tag{4.18}$$

Then for $k < \min(1/\varepsilon, 1/(1-\varepsilon))$, we have $d_1 < r < d_2$. However, if $\varepsilon < 1/2$, $d_2 \to \infty$ as $k \to 1/(1-\varepsilon)$ while $d_1 \to r(1-2\varepsilon)/2(1-\varepsilon)$. Hence, if $d_1' \leq r(1-2\varepsilon)/2(1-\varepsilon)$ we can only censor η_1 for negative values. In a similar fashion, it can be shown that if $\varepsilon > 1/2$ and $d_2' \geq 2\varepsilon r/(2\varepsilon - 1)$ we can censor only positive values of η_1.

Theorems about symmetric censoring

Lemma 1. *Let*

$$d_{1m} = \max\left[0, \frac{r}{2}\frac{1-2\varepsilon}{1-\varepsilon}\right]$$

and

$$d_{2m} = \left[\max\left[0, \frac{2\varepsilon-1}{2\varepsilon r}\right]\right]^{-1}$$

where 1/0 is defined to be infinity. If d_2 is given by (4.11), $d_{1m} < d_1 \leq d_1'$ and $d_2' \leq d_2 < d_{2m}$, then there exist c_1 and c_2, at least one of which is greater than or equal to unity, which solve (4.13) and (4.14).

Proof. See [8].

Theorem 1. *For any pair (d_1, d_2) which satisfies the conditions of the lemma let $p(d_1) = f(d_1') - f(d_1)$ and $q(d_2) = g(d_2') - g(d_2)$. A necessary and sufficient condition for both c_1 and c_2 to equal or exceed unity is $d_1q \leq p \leq d_2q$.*

Proof. See [8].

Corollary. *A sufficient condition for the existence of a set A with $c_i > 1$ is that there is some pair (d_1, d_2), $d_1 < d_2'$, which satisfies the conditions of the lemma and for which $f(d_1)/g(d_2) = r$.*

Proof. See [8].

The advantage to stating the conditions for the existence of a solution in terms of $f(d_1)$ and $g(d_2)$ is that these functions can be tabulated and stored for any pair of distributions μ_{0e} and μ_{1e}. We have done so for the case where μ_{0e} is $N(1,1)$ and μ_{1e} is $N(0,1)$. Figures 1-4 show the regions in the rectangle $0 < \alpha_0 < 0.8$, $0 < \alpha_1 < 0.95$ for which a set A can be found for which μ_{0c}' and μ_{1c}' are density functions. The values of ε in Figs. 1-4 are 0, 0.15, 0.3 and 0.45, respectively. It should be noted that as $\varepsilon \to 1/2$ the region becomes larger. This appears to be due to the fact that d_{1m} decreases and as a result there is a greater range of possible values for d_1.

Theorem 2. *For any pair (d_1, d_2) which satisfies the conditions of the lemma a sufficient condition for both c_1 and c_2 to equal or exceed unity is $\alpha_1 \geq \max(r/d_2, 1-r+rd_1)$. For $\varepsilon < 1/2$, the sufficient condition will be satisfied for some (d_1, d_2) if and only if $\alpha_0 > d_{1m}/r$. For $\varepsilon > 1/2$ the sufficient condition will be satisfied for some (d_1, d_2) if and only if $\alpha_1 > rd_{2m}^{-1}$.*

Proof. See [8].

A major problem with symmetric censoring is that the relationship between d_1 and d_2 depends on the unknown value of ε. This is especially troubling since this relationship must be exact so that both μ_{0h} and μ_{1h} are confined to the set on which the nonlinearity has max-

imum mass and minimum slope. It is important to note, however, that there is no real necessity to know the values of d_1 and d_2 in order to use the robust estimator. This is due to the fact that p and q are continuous in d_1 and d_2, respectively. Hence, if there is a pair (d_1,d_2) which satisfies the Theorem, it is not, in general, unique. As a result, slight changes in the value of ε which result in slight perturbations in the relationship between d_1 and d_2 (see (4.11)) will cause only slight changes in the values of c_1 and c_2. As can be seen from Eqs. (4.13)-(4.14) and (4.17)-(4.18) the value of k completely determines the set A. Hence, given α_0 and α_1, there is, in general, some value of k for which A includes $c_i \geq 1$, i=1,2, over a range of values of ε. We need not bother to find d_1 and d_2 if we are certain that the actual value of ε falls within that range.

Theorems about one-sided censoring

Theorem 3. *A necessary and sufficient condition for one-sided censoring at a negative value of k is that there exists some d_1, $0 < d_1 < d_1^* = \min(d_1', d_{1m})$ such that $\alpha_1 = p(d_1)/(p(d_1) + d_1)$. If $d_1' = d_1^*$ and $\varepsilon < 1/2$ such a solution always exists. The relationship between d_1 and k is given by (4.17).*

Proof. See [8].

Theorem 4. *A necessary and sufficient condition for one-sided censoring at a positive value of k is that there exists some d_2, $\max(d_2', d_{2m}) = d_2^* < d_2 < \infty$ such that $\alpha_0 = d_2 q(d_2)/(d_2 q(d_2) + 1)$. If $d_2' = d_2^*$ and $\varepsilon > 1/2$ such a solution always exists. The relationship between d_2 and k is given by (4.18).*

The proof of Theorem 4 is virtually identical to that of Theorem 3.

§5 Conclusions

In this paper we investigated the problem of estimating a mixture constant from observations of mixture noise. In particular the properties of M-esimators for ε were discussed. It was shown that estimators of this type are, within certain mild restrictions, consistent and asymptotically normal and that the asymptotic variance is minimized by the maximum likelihood estimator. The MLE was discussed in detail as were M-estimators based on the error between the sample moments and the moments given by the estimate of ε. Both were found to be consistent and asymptotically normal. The remainder of the chapter contains the formulation and solution of the robust estimation problem for the case where some uncertainty exists about the nominal and contaminating distributions. The robust test for detecting the presence of vanishingly small contaminants was also found.

Acknowledgements

This research was supported by the Office of Naval Research under Grant N00014-81-K-0146 and by the National Science Foundation under Grant ECS83-17777.

References

[1] P.J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, pp. 73-101, Mar. 1964.

[2] ----, "A robust version of the probability ratio test," *Ann. Math. Statist.*, vol. 36, pp. 1753-1758, Dec. 1965.

[3] K.L. Chung, *A Course in Probability Theory*, 2nd ed., Academic Press, New York, 1974.

[4] H.L. Royden, *Real Analysis*, 2nd ed., Macmillan Publishing Co., New York, 1968.

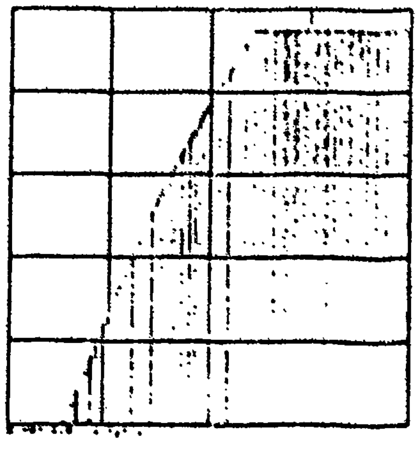
[5] R.D. Martin and S.C. Schwartz, "Robust detection of a known signal in nearly Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 50-56, Jan. 1971.

[6] G.R. Wilson and D.R. Powell, "Experimental and modeled density estimates of underwater acoustic returns," *Proceedings of Signal Processing in the Ocean Environment Workshop*, Annapolis, MD, May 1982.

[7] R.E. Quandt and J.B. Ramsey, "Estimating mixtures of normal distributions and switching regressions," *J. of American Statistical Ass.*, vol. 73, no. 364, pp. 730-752, December 1978.

[8] D.J. Warren, "Detection and estimation in the presence of uncertainty," *PhD Dissertation*, Dept. of Electrical Engineering, Princeton University, Princeton, NJ, 1987.

α_1



α_0

Fig. [illegible]

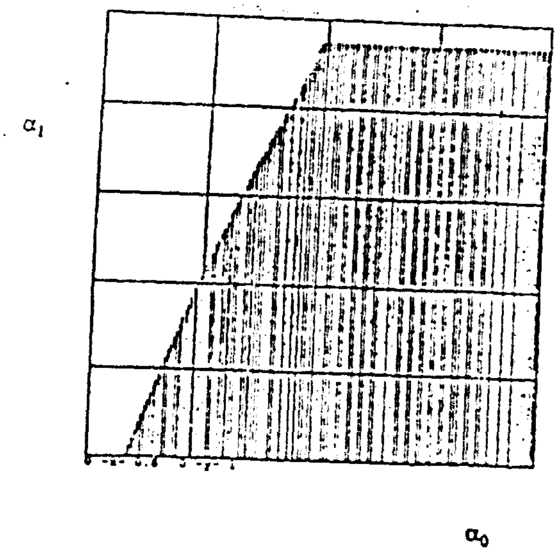


Fig. 2 Region of solutions, α_1 vs. α_0, $\epsilon = 0.15$. μ_0 is $N(1,1)$, μ_1 is $N(0,1)$

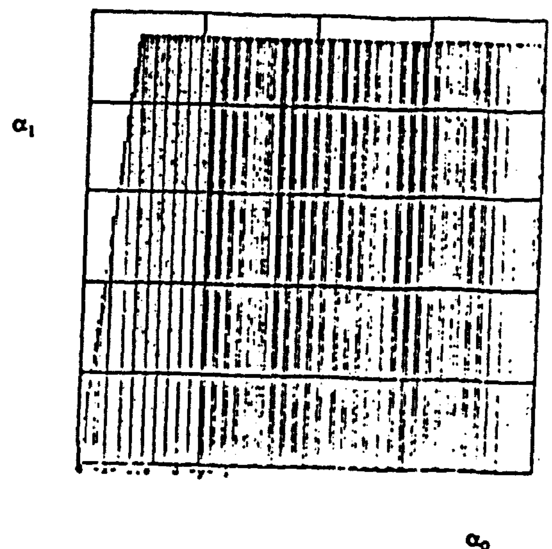


Fig. 4 Region of solutions, α_1 vs. α_0, $\epsilon = 0.45$. μ_0 is $N(1,1)$, μ_1 is $N(0,1)$

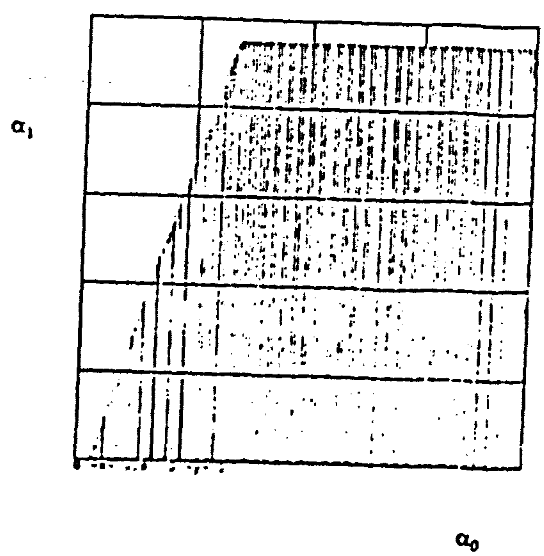


Fig. 3 Region of solutions, α_1 vs. α_0, $\epsilon = 0.3$. μ_0 is $N(1,1)$, μ_1 is $N(0,1)$